

AD-A082 239

CALIFORNIA UNIV SAN DEIGO LA JOLLA DEPT OF ELECTRICAL--ETC F/6 9/3
CONSISTENT ESTIMATION OF CONTINUOUS-TIME SIGNALS FROM NONLINEAR--ETC(U)
MAR 80 E MASRY, S CAMBANIS N00014-75-C-0652

UNCLASSIFIED

NL

[]

ADP 159

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

[]

END
DATE
FILMED
4 80
DTIC

Unclassified, March 10, 1980

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (cont.)

distribution. These rather unexpected results hold for all bounded and uniformly continuous signals. In addition to the hardlimiter, such results are also established for certain monotonic and non-monotonic nonlinearities.

S/N 0102- LF-014-6601

Unclassified, March 10, 1980

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONSISTENT ESTIMATION OF CONTINUOUS-TIME SIGNALS
FROM NONLINEAR TRANSFORMATIONS OF NOISY SAMPLES

by
Elias Masry and Stamatis Cambanis

Abstract

In general, a signal cannot be reconstructed from its sign, i.e., from its hardlimited version. However, by deliberately adding noise to samples of the signal prior to hardlimiting, it is shown that the signal can be estimated consistently from its hardlimited noisy samples as the sampling rate tends to infinity. In fact, such estimates are shown to converge with probability one to the signal and also, to be asymptotically normal. The estimates, which are generally nonlinear, can be made linear by a proper choice of the noise distribution. These rather unexpected results hold for all bounded and uniformly continuous signals. In addition to the hardlimiter, such results are also established for certain monotonic and non-monotonic nonlinearities.

The work of E. Masry was supported by the Office of Naval Research under Contract N00014-75-C-0652. The work of S. Cambanis was supported by the Air Force Office of Scientific Research under Grant AFOSR-75-2796.

E. Masry is with the Electrical Engineering & Computer Sciences Department, University of California at San Diego, La Jolla, CA 92093.

S. Cambanis is with the Department of Statistics, University of North Carolina, Chapel Hill, NC 27514.

Accession For	
NTIS Grant	<input checked="checked" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	<input type="checkbox"/>
By	
Date	
Distribution Codes	
Dist	Special

80 3 24 008

I. INTRODUCTION

In this paper we study the problem of reconstructing a real signal $s(t)$ defined on an interval I , from certain nonlinear transformations of its samples $\{s(k/W)\}_k$ that are deliberately corrupted by additive noise $\{X_k\}_k$, i.e., from $\{f[s(k/W) + X_k]\}_k$ where $f(x)$ is a memoryless nonlinearity such as a hardlimiter. Under appropriate conditions it is shown that a properly chosen, generally nonlinear, estimate $\hat{s}_W(t)$ of $s(t)$ converges in quadratic mean, as well as with probability one, to $s(t)$ as the sampling rate W tends to infinity. It should be pointed out that the memoryless nonlinearity $f(x)$ need not be one-to-one so that the signal $s(t)$ cannot, in general, be recovered from $\{f[s(k/W)]\}_k$ as W tends to infinity, in the absence of the additive noise $\{X_k\}$. It is the deliberate addition of the noise that makes the reconstruction of the signal feasible.

This work is motivated by the observation that an arbitrary continuous function $s(t)$, $-\infty < t < \infty$ cannot, in general, be reconstructed from its sign, $\text{sgn}[s(t)]$, $-\infty < t < \infty$. This situation remains true even when the function $s(t)$ is analytic, e.g., bandlimited. We recall that for a bandlimited function $s(t) = \int_{-W}^W e^{it\lambda} S(\lambda) d\lambda$, $S(\lambda) \in L_1[-W, W]$, we have by Titchmarsh's theorem [1] the conditionally convergent product $s(t) = s(0) \prod_{n=1}^{\infty} (1 - t/z_n)$ where $s(0) \neq 0$ and $\{z_n\}$ is the set of all (real and complex) zeros of $s(z)$, $z = t + iu$, in the complex plane. Thus $s(t)$ cannot be determined from its zero crossings since the complex zeros are not observable. Duffin and Schaeffer [2] have shown that the function $r(z) \triangleq C \cos Wz - s(z)$, $C > \sup_t |s(t)|$, has real simple zeros $\{t_n\}$ only and $r(t) = r(0) \prod_{n=1}^{\infty} (1 - t/t_n)$ so that

$$s(t) = C \cos Wt - [C - s(0)] \prod_{n=1}^{\infty} \left(1 - \frac{t}{t_n}\right).$$

Hence, $s(t)$ can be determined by the zero crossings of $C \cos Wt - s(t)$. This result has found no practical use in communication systems since the identification of the

zero crossing points $\{t_n\}$ of $C \cos \omega t - s(t)$ as well as the formation of the infinite product $\prod_n (1 - t/t_n)$ are not easily implemented. More significantly, no digital reconstruction scheme of $s(t)$, based on samples of $\text{sgn}[C \cos \omega t - s(t)]$, is available.

It will be shown in Section II that for all bounded uniformly continuous signals $s(t)$ (not necessarily bandlimited) we have estimates $\hat{s}_W(t)$ of $s(t)$, based on the binary data $\{\text{sgn}[s(k/W) + X_k]\}_k$, which converge with probability one to $s(t)$, as the sampling rate W tends to infinity. It is the deliberate corruption of the samples of the signal by the noise, before hardlimiting, that makes it possible to reconstruct $s(t)$ from the output of the hardlimiter. Moreover, by properly choosing the distribution of the noise, we can make the estimate to be linear.

The general problem can be modelled as a transmitter/receiver (with no channel noise) with a structure depicted in Figure 1. A continuous-time signal $s(t)$ on an interval I is sampled at equally-spaced points $\{k/W\}_k$ in I where W is the sampling rate. The samples $\{s(k/W)\}_k$ are then deliberately corrupted by additive noise $\{X_k\}_k$ which is a sequence of independent identically distributed random variables whose distribution is specified below. The noisy samples $\{Y_{W,k} \triangleq s(k/W) + X_k\}_k$ are passed through a given memoryless nonlinearity $f(x)$ which need not be monotonic, a typical example being a hardlimiter. Its output sequence $\{Z_{W,k} \triangleq f(Y_{W,k})\}_k$ is transmitted. The receiver structure is generally nonlinear and consists of a linear system $h_W = \{h_W(t, k)\}_k$ cascaded with a memoryless nonlinearity $g(x)$. The output $\hat{m}_W(t)$ of the linear system h_W is given by

$$\hat{m}_W(t) = \sum_k Z_{W,k} h_W(t, k), \quad t \in I. \quad (1)$$

The choice of the linear system does not depend on the signal $s(t)$ nor on the distribution of the noise nor on the nonlinearity $f(x)$ in the transmitter; it only depends on the time interval I . On the other hand, the nonlinearity $g(x)$ in the receiver is determined by the distribution of the noise and the nonlinearity $f(x)$ in the transmitter. The estimate $\hat{s}_W(t)$ of $s(t)$ is defined by

$$\hat{s}_W(t) = g[\hat{m}_W(t)], \quad t \in I. \quad (2)$$

The main results of the paper are the mean-square consistency of the estimate (2) (Theorems 2.1 and 3.1), its strong consistency (Theorems 2.3 and 3.3), and a central limit theorem for the error $\hat{s}_W(t) - s(t)$ (Theorems 2.4 and 3.4). Of possible independent interest are the convergence properties (Theorems 4.0-4.1) of $\hat{m}_W(t)$ as an estimate of the mean function $m(t) = E[f(s(t)+X)]$ of the output of the nonlinearity f .

The feasibility of the reconstruction of the signal was suggested by the results of a recent paper [3] by the authors; according to which $s(t)$ can be determined from the mean function $m(t) = E[f(s(t)+X)]$. This suggests that an estimate of $s(t)$ can be obtained from an estimate of $m(t)$ via (2). The form of the estimate (1) for $m(t)$, i.e., the linear system in the receiver, was motivated by the work of Dorogovcev [4] on the nonparametric estimation of regression functions.

Throughout the manuscript we shall assume that $s(t)$ belongs to the following class of signals.

Assumption A. Let b be a fixed known positive constant, and $s(t)$ be any uniformly continuous function on the interval I (finite or infinite) satisfying $|s(t)| \leq b$ for all $t \in I$.

As a consequence, the receiver structure and the convergence results of this paper are nonparametric in the signal. Incidentally, additional assumptions on the signal, such as differentiability or bandlimitness, do not provide an improvement in the rate of convergence.

The organization of the paper is as follows: Due to its apparent practical significance, the case of a hardlimiter, $f(x) = \text{sgn } x$, is presented and discussed separately in Section II. The general case is considered in Section III. In Section IV the convergence properties of the estimate $\hat{m}_W(t)$ are obtained. The derivations of the theorems stated in Sections II and III are given in Section V.

Throughout this paper, the expressions $o(\cdot)$ and $O(\cdot)$ as $W \rightarrow \infty$ are uniform in t over closed and bounded intervals in the interior of the interval I . This property will not be repeated in the statements of the theorems.

II. THE HARDLIMITER CASE

In this section we consider the hardlimiter case, $f(x) = \text{sgn } x$, for which the transmitted data is binary. We make the following assumptions.

(i) The signal $s(t)$ satisfies assumption A and the interval I is either $[0,1]$ or $[0,\infty)$ (other choices are discussed in Section III).

(ii) The distribution of the noise X is either normal with mean zero, known variance σ^2 and density $\phi(x;\sigma)$, or uniform over $[-b,b]$. (Other appropriate distributions, such as Laplacian, could also be used.)

Define the function $\mu(s)$ by

$$\mu(s) = E[\text{sgn}(s+X)], \quad -\infty < s < \infty.$$

When X is normal, $\mu(s)$ is given by

$$\mu_N(s) = \sqrt{2/\pi} \int_0^{s/\sigma} e^{-u^2/2} du, \quad -\infty < s < \infty, \quad (3a)$$

and when X is uniform over $[-b,b]$, we have

$$\mu_U(s) = \begin{cases} -1 & , \quad s < -b \\ s/b & , \quad -b \leq s \leq b \\ 1 & , \quad b < s \end{cases} \quad (3b)$$

Note that $\mu_N(s)$ and $\mu_U(s)$ are strictly monotonic over $(-\infty, \infty)$ and $[-b,b]$, respectively.

We now specify the structure of the receiver. When X is normal, the nonlinearity $g(x)$ is chosen as

$$g_N(x) = \begin{cases} \mu_N^{-1}(x), & |x| \leq \mu_N(c) \\ 0 & , \quad |x| > \mu_N(c) \end{cases}, \quad c = b + \epsilon, \quad \epsilon > 0. \quad (4a)$$

When X is uniform over $[-b,b]$, the nonlinearity $g(x)$ is chosen as

$$g_U(x) = \begin{cases} bx & , \quad |x| \leq 1 \\ 0 & , \quad |x| > 1 \end{cases}. \quad (4b)$$

The choice of the linear system $h_W = \{h_W(t, k)\}_k$ depends only on the interval I .

When $I = [0, \infty)$, h_W is defined by

$$h_W(t, k) = \frac{(Wt)^k}{k!} e^{-Wt}, \quad k = 0, 1, \dots, t \geq 0, W > 0, \quad (5)$$

and when $I = [0, 1]$, by

$$h_W(t, k) = \binom{W}{k} t^k (1-t)^{W-k}, \quad k = 0, 1, \dots, W, 0 \leq t \leq 1, \quad (6)$$

W : positive integer.

A more general class of linear systems is considered in Section III. With

$$\hat{m}_W(t) = \sum_k \operatorname{sgn} \left[s\left(\frac{k}{W}\right) + x_k \right] h_W(t, k), \quad t \in I, \quad (7)$$

representing the output of either linear system (5) or (6), the estimate $\hat{s}_W(t)$ of $s(t)$ is given by

$$\hat{s}_W(t) = g_N[\hat{m}_W(t)], \quad t \in I \quad (8a)$$

when X is normal, and by

$$\hat{s}_W(t) = b \hat{m}_W(t), \quad t \in I \quad (8b)$$

when X is uniform over $[-b, b]$. (Since $|\hat{m}_W(t)| \leq 1$, only the linear portion of $g_U(x)$ is used.) Thus in the latter case, the estimate $\hat{s}_W(t)$ is linear in the data $\{\operatorname{sgn}[s(k/W) + x_k]\}_k$.

Our first result shows the mean-square consistency of the estimate $\hat{s}_W(t)$ and provides bounds on the rate of convergence; it is stated in terms of the modulus of continuity of $s(t)$ defined by

$$\omega(s; \delta) = \sup_{\{t, t' \in I: |t-t'| < \delta\}} |s(t) - s(t')|, \quad \delta > 0.$$

Theorem 2.1. (a) If $I = [0, 1]$ and the linear system is determined by (6) then for every $0 < t < 1$, the estimates $\hat{s}_W(t)$, given by (8a) and (8b), converge in the mean-square sense to $s(t)$, as $W \rightarrow \infty$, and

$$E[\hat{s}_W(t) - s(t)]^2 \leq K_1 \omega^2(s; \sqrt{t(1-t)/W}) + K_2 \frac{1 + o(1)}{2\sqrt{\pi W t(1-t)}}. \quad (9a)$$

(b) if $I = [0, \infty)$ and the linear system is determined by (5) then for every $t > 0$, the estimates $\hat{s}_W(t)$, given by (8a) and (8b), converge in the mean-square sense to $s(t)$, as $W \rightarrow \infty$, and

$$E[\hat{s}_W(t) - s(t)]^2 \leq K_1 \omega^2(s; \sqrt{t/W}) + K_2 e^{-2Wt} I_0(2Wt), \quad (9b)$$

where $I_0(x)$ is the modified Bessel function of the 1st kind of order zero and $\exp(-2Wt)I_0(2Wt) = [1 + o(1)]/\sqrt{4\pi Wt}$.

The constants K_1, K_2 are the same for both parts (a) and (b) and are given as follows for the estimates (8a) and (8b).

$$\text{For (8a): } K_1 = \frac{8}{\pi \sigma^2} K_2, \quad K_2 = \frac{1 + (b/\epsilon)^2}{4\phi^2(b + \epsilon; \sigma)},$$

$$\text{For (8b): } K_1 = 4, \quad K_2 = b^2.$$

In the bounds (9) on the mean-square error, the first term is due to the bias of the estimate whereas the second term is due to its variance; the bias depends on the modulus of continuity of the signal whereas the variance is always $o(W^{-1/2})$. For example, if $s(t)$ is Lip γ , $0 < \gamma \leq 1$, then the mean-square error is $o(W^{-\min(\gamma, 1/2)})$, and for $\gamma > 1/2$, it is dominated by the variance and is $o(W^{-1/2})$. Additional smoothness conditions on the signal $s(t)$, such as differentiability, would provide faster convergence rate for the bias but would not improve the rate of convergence of the mean-square error.

When X is normal, the constants K_1 and K_2 depend on the variance σ^2 of X and on ϵ (cf. (4a)). When the variance is asymptotically dominant, e.g. if the signal is Lip γ with $1/2 < \gamma \leq 1$, asymptotically optimal choices for σ and ϵ can be

found by minimizing K_2 ; we find $\sigma = 2b$ and $\epsilon = b$ for which $K_2 = 4\pi eb^2$ and $K_1 = 8e$ (and these values are larger than those when X is uniform).

The next theorem shows that the estimates $\hat{s}_W(t)$ converge to $s(t)$ in the $2\ell^{\text{th}}$ mean for every integer $\ell \geq 1$ and that faster rates of convergence are available in this case.

Theorem 2.2. Let $s(t)$ be Lip γ on I , $0 < \gamma \leq 1$. Then for all t in the interior of I and for every integer $\ell \geq 1$ the estimates (8a) - (8b) satisfy

$$E[\hat{s}_W(t) - s(t)]^2 = o(W^{-\ell \min(\gamma, 1/2)}) .$$

From the practical point of view, convergence of the estimate $\hat{s}_W(t)$ to $s(t)$ with probability one (rather than in the mean) is preferable so that $s(t)$ can be reconstructed from almost every realization of the data $\{\text{sgn}[s(k/W) + X_k]\}_k$, i.e., corresponding to almost every realization of the noise sequence $\{X_k\}_k$. This strong consistency of the estimate $\hat{s}_W(t)$, along with its rate of convergence, is given in the next theorem.

Theorem 2.3. Let $s(t)$ be Lip γ on I , $0 < \gamma \leq 1$, and let α be any constant satisfying $0 < \alpha < (1/2)\min(\gamma, 1/2)$.

(a) If the linear system is determined by (5) then for each fixed $t \in (0, \infty)$ and each fixed sequence $W_n \uparrow \infty$ as $n \uparrow \infty$, we have with probability one

$$(W_n)^\alpha |\hat{s}_{W_n}(t) - s(t)| \rightarrow 0 \text{ as } n \rightarrow \infty ,$$

(b) If the linear system is determined by (6), then for each fixed $t \in (0, 1)$ and with $W \equiv n$, a positive integer, we have with probability one

$$n^\alpha |\hat{s}_n(t) - s(t)| \rightarrow 0 \text{ as } n \rightarrow \infty .$$

Our final result in this section provides a central limit theorem for the error $\hat{s}_W(t) - s(t)$. When the noise X is uniform, we shall assume that it is uniform

over $[-c, c]$ with $c > b$, in which case the estimate (8b) is replaced by $\hat{s}_W(t) = c \hat{m}_W(t)$.

Theorem 2.4. Let $s(t)$ be Lip γ on I , $1/2 < \gamma \leq 1$. Define

$$\tilde{s}_W(t) = \beta_W(t) [\hat{s}_W(t) - s(t)], \quad t \in I,$$

where the normalizing factor $\beta_W(t)$ is specified below.

(a) If $I = [0, \infty)$ and the linear system is determined by (5) then the values of the process $\{\tilde{s}_W(t), 0 < t < \infty\}$ at distinct instants $\{t_i\}$ are asymptotically independent standard normal variables as $W \rightarrow \infty$.

(b) If $I = [0, 1]$ and the linear system is determined by (6) then for each fixed $t \in (0, 1)$, $\tilde{s}_W(t)$ is asymptotically standard normal variable as $W \rightarrow \infty$.

When X is uniform over $[-c, c]$, $\beta_W(t)$ is given by

$$\beta_W(t) = \{c^2 \text{Var}[\hat{m}_W(t)]\}^{-1/2},$$

and when X is normal, by

$$\beta_W(t) = 2 \phi[s(t), \sigma] \{\text{Var}[\hat{m}_W(t)]\}^{-1/2}.$$

Upper and lower bounds on $\beta_W(t)$ can be obtained from the bounds (35) on $\text{Var}[\hat{m}_W(t)]$.

The asymptotic normality of $\tilde{s}_W(t)$ can be used to obtain confidence intervals for the error $\hat{s}_W(t) - s(t)$ by using the bounds on $\beta_W(t)$.

We conclude this section with some practical comments on the various transmitter/receiver combinations. Clearly, the simplest transmitter uses uniformly distributed noise and the corresponding receiver is then linear. The "Bernstein" linear receiver would be the simplest to use since it employs a finite number $(W+1)$ of samples to reconstruct the signal over the interval $[0, 1]$. The actual sampling rate W to be used can be determined from Theorem 2.1 to correspond to an acceptable mean-square error. For signals defined over $[0, \infty)$, aside from using the "Szász" linear receiver, one could also use the "Bernstein" linear receiver sequentially over consecutive intervals of unit length.

III. THE GENERAL CASE

In this section we consider general (nonconstant) nonlinearities $f(x)$ in the transmitter and, under appropriate conditions on $f(x)$, we specify noise distributions, linear systems h_W , and memoryless nonlinearities $g(x)$ such that $\hat{s}_W(t)$, defined by (2) and (1), is a consistent estimate of the signal $s(t)$ as $W \rightarrow \infty$. Theorems 3.1-3.4 contain Theorems 2.1-2.4 as special cases.

We first specify the distribution of the noise X , introduce appropriate assumptions on $f(x)$, and specify the memoryless nonlinearity $g(x)$ in the receiver. There are two types of symmetric distributions appropriate here, those supported by the entire real line $(-\infty, \infty)$, and those supported by the finite interval $[-b, b]$. For the sake of concreteness we will concentrate on two such typical distributions, the normal $N(0, \sigma^2)$ with density $\phi(x; \sigma)$ and the uniform over $[-b, b]$. We shall use the function $\mu(s)$ defined by

$$\mu(s) = E[f(s+X)], \quad -\infty < s < \infty \quad (10)$$

Clearly, $\mu(s)$ depends on f and on the distribution of X . When X is $N(0, \sigma^2)$, we have

$$\mu_N(s) = \int_{-\infty}^{\infty} f(s+x) \phi(x; \sigma) dx,$$

and when X is uniform over $[-b, b]$ we have

$$\mu_U(s) = \frac{1}{2b} \int_{-b}^b f(s+x) dx.$$

In the particular case when $f(x) = \text{sgn } x$, $\mu_U(s)$ and $\mu_N(s)$ are given by (3). For monotonic nonlinearities $f(x)$ (which need not be strictly monotonic, e.g., the hardlimiter), and for the case of nonlinearities $f(x)$ described in (B2) below (which need not be monotonic, e.g. $f(x) = x^3 - \sigma^2 x$), it has been shown in [3] that $\mu_N(s)$ is strictly monotonic, and thus its inverse $\mu_N^{-1}(x)$ exists. We now specify the memoryless nonlinearity $g(x)$ in the receiver, for various classes of transmitter-nonlinearities $f(x)$ and noise distributions.

Assumption B. We say that (B) is satisfied if any one of (B1), (B2), or (B3) is satisfied.

(B1): i. X is $N(0, \sigma^2)$.
 ii. $f(x) \in L_4[\phi(x; \sigma)dx]$ and is monotonic (not necessarily strictly).

iii. $g(x) = \begin{cases} \mu_N^{-1}(x), & \mu(-c) \leq x \leq \mu(c) \\ 0, & \text{otherwise} \end{cases}, c = b + \epsilon, \epsilon > 0.$

(B2): i. X is $N(0, \sigma^2)$.
 ii. $f(x) \in L_4[\phi(x; \sigma)dx]$ is an odd function and has nonnegative Hermite coefficients $\{e_k\}_k$ with $e_1 > 0$. (See [3].)

iii. $g(x) = \mu_N^{-1}(x)$ for $-\infty < x < \infty$.

(B3): i. X is uniform over $[-b, b]$.

ii. $f(x) = \text{sgn } x$.

iii. $g(x) = \begin{cases} bx, & |x| \leq 1 \\ 0, & |x| > 1. \end{cases}$

Our first result shows the mean-square consistency of the estimate $\hat{s}_W(t)$ under general conditions on the linear system $h_W = \{h_W(t, k)\}_k$.

Theorem 3.0. Let Assumptions (A) and (B) be satisfied. For every $t \in I$ for which

- i. $h_W(t, k) \geq 0$, for all k ,
- ii. $\sum_k h_W(t, k) = 1$,
- iii. $\sum_k \left(t - \frac{k}{W}\right)^2 h_W(t, k) \rightarrow 0$ as $W \rightarrow \infty$,
- iv. $\sum_k h_W^2(t, k) \rightarrow 0$ as $W \rightarrow \infty$,

the estimate (2) converges in quadratic mean to $s(t)$ as $W \rightarrow \infty$.

The first condition on h_W makes the linear system a positive linear operator, the second is a summability/normalization condition, the third guarantees

that the bias of the estimate tends to zero, and the fourth condition guarantees that the variance of the estimate goes to zero. A large class of linear systems h_W satisfying the conditions of Theorem 3.0 can be obtained as follows.

Proposition 3.0. Let $\{\xi_i\}_{i=1}^{\infty}$ be a sequence of independent identically distributed random variables with integer values, mean $t \in I$, and finite second moments. Then $\{h_n(t, k)\}_k$ defined

$$h_n(t, k) = \Pr\{\xi_1 + \dots + \xi_n = k\}, \quad n = 1, 2, \dots, k=0, \pm 1, \dots \quad (11)$$

satisfies assumptions (i)-(iv) of Theorem 3.0 (with W taking positive integer values) for every $t \in I$ for which the random variable ξ_i is not degenerate.

Positive linear operators of the type described in Proposition 3.0 have been considered in the approximation theory literature [5], where conditions (i)-(iii) of Theorem 3.0 are established and used for the interpolation of continuous functions $m(t)$ on I by $\sum_k m(k/n)h_n(t, k)$. We mention two examples: When each ξ_i takes on the values 0 and 1 with $\Pr\{\xi_i = 1\} = t$, then h_n is given by (6) and represents the Bernstein operator. When each ξ_i is Poisson with parameter t , then h_n is given by (5) (with $W = n$) and represents the Szász operator.

Theorem 3.0, while guaranteeing mean-square consistency of the estimate $\hat{s}_W(t)$, provides no bounds on the rate of convergence. We shall derive such bounds for linear systems h_W corresponding to the class of generalized Szász operators [6] (see below) and to the Bernstein operator. While the Szász operator (5) can be generated as in Proposition 3.0, the class of generalized Szász operators cannot. We consider the entire class of generalized Szász operators, rather than the single Szász operator, because with no additional work we obtain the same rates of convergence for this entire class of linear systems.

We now introduce the generalized Szász operators. Let $A(z) = \sum_{n=0}^{\infty} a_n z^n$ be an analytic function in $|z| < R$, for some $R > 1$, and suppose that for all n

$$a_n \geq 0 \text{ and } A(1) = \sum_{n=0}^{\infty} a_n > 0 .$$

The Appel polynomials [7] $p_k(u)$, $u \geq 0$, are defined by their generating function

$$A(z) e^{uz} = \sum_{k=0}^{\infty} p_k(u) z^k , \quad (12)$$

i.e.,

$$p_k(u) = \sum_{j=0}^k a_{k-j} \frac{u^j}{j!} .$$

The generalized Szász operator is represented by $h_W = \{h_W(t, k)\}_k$ with

$$h_W(t, k) = p_k(Wt) \frac{e^{-Wt}}{A(1)} , \quad k = 0, 1, \dots, t \geq 0, W > 0 . \quad (13)$$

The Szász operator corresponds to $A(z) \equiv 1$ for which $p_k(u) = u^k/k!$.

The following assumption specifies the interval I and the linear system h_W .

Assumption C. We say that (C) is satisfied if either (C1) or (C2) is satisfied.

(C1): $I = [0, \infty)$ and h_W is a generalized Szász operator defined by (13) and (12).

(C2): $I = [0, 1]$ and h_W is the Bernstein operator defined by (6).

We shall therefore concentrate on signals $s(t)$ defined on the positive real line $[0, \infty)$ or the unit interval $[0, 1]$. By appropriate scaling one can similarly consider signals defined on any half line or any finite interval. The case of signals defined on the entire real line can be reduced to the positive real line by separately considering the parts of the signal on $[0, \infty)$ and $(-\infty, 0]$.

All the following results hold under Assumptions (A), (B) and (C).

Assumption (A) states the conditions on the signal $s(t)$, Assumption (B) determines the nonlinearity $g(x)$ in the receiver, and Assumption (C) determines the linear system h_W in the receiver.

Our next result provides upper bounds on the mean-square estimation error.

Theorem 3.1. Under Assumptions (A), (B) and (C) we have for each $t \in I$, that the estimate (2) satisfies

$$E[\hat{s}_W(t) - s(t)]^2 \leq K_1 \omega^2(s; \alpha_W(t)) + K_2 v_W^2(t)$$

where $\alpha_W^2(t)$ and $v_W^2(t)$ are determined by (C),

$$\text{for (C1): } \alpha_W^2(t) = \frac{t}{W} + \frac{A(1) + A'(1)}{A(1)W^2}, \quad v_W^2(t) = \frac{1 + o(1)}{2\sqrt{\pi Wt}},$$

$$\text{for (C2): } \alpha_W^2(t) = \frac{t(1-t)}{W}, \quad v_W^2(t) = \frac{1 + o(1)}{2\sqrt{\pi Wt(1-t)}},$$

and the constants K_1 and K_2 are determined by (B),

$$\text{for (B1): } K_1 = 4Q^2(q^{-2} + (b/\Delta)^2), \quad K_2 = U_2(q^{-2} + (b/\Delta)^2)$$

$$\text{for (B2): } K_1 = 4Q^2/q^2, \quad K_2 = U_2/q^2,$$

$$\text{for (B3): } K_1 = 4b^2Q^2, \quad K_2 = b^2U_2,$$

and the constants q , Q , U_2 and Δ are defined in (17).

It follows that $\hat{s}_W(t)$ converges to $s(t)$ in quadratic mean for every t in the interior of the interval I , i.e., for $t > 0$ under (C1) and $0 < t < 1$ under (C2). Also, for the entire class of generalized Szász operators, the rate of decay of $\alpha_W^2(t)$ and $v_W^2(t)$ as $W \rightarrow \infty$ is the same, $O(1/W)$ and $O(1/\sqrt{W})$, respectively, and thus the rate of convergence of the bound on the mean-square error is also the same. This rate is also identical to that of the Bernstein operator. For example, when $s(t) \in \text{Lip } \gamma$, $0 < \gamma \leq 1$, the mean-square error is $O(W^{-\min(\gamma, 1/2)})$ for all choices of linear and nonlinear systems h_W and $g(x)$, covered by Theorem 3.1.

Bounds on the higher order moments of the estimation error can be obtained in a similar manner and they provide faster rates of convergence.

Theorem 3.2. Assume that $s(t)$ is Lip γ on I , $0 < \gamma \leq 1$, and that Assumptions (A), (B) and (C) are satisfied. Let ℓ be a positive integer and under (B1) or (B2) assume, in addition, that $f(x) \in L_{2\ell}[\phi(x;\sigma)dx]$. Then for every t in the interior of I , the estimate (2) converges in the $2\ell^{\text{th}}$ mean to $s(t)$ as $W \rightarrow \infty$ and for some continuous function $K_{\ell,\gamma}(t)$,

$$E[\hat{s}_W(t) - s(t)]^{2\ell} \leq \frac{K_{\ell,\gamma}(t)}{W^{\ell \min(\gamma, 1/2)}} [1 + o(1)] .$$

The exact expression for $K_{\ell,\gamma}(t)$ is quite involved but easily expressed in terms of $F_{\ell,\gamma}(t)$, introduced in the proof of Theorem 4.2, and the constants in Proposition 5.1. The bound of Theorem 3.2 can be used to obtain the strong consistency of the estimate $\hat{s}_W(t)$ and the rate of almost sure convergence.

Theorem 3.3. Assume that $s(t)$ is Lip γ on I , $0 < \gamma \leq 1$, that Assumptions (A), (B) and (C) are satisfied, and in the case of (B1) or (B2) that $f(x) \in L_{2\ell}[\phi(x;\sigma)dx]$ for some positive integer ℓ satisfying $\ell \geq 1 + \gamma^{-1}$ for $0 < \gamma < 1/2$, and $\ell \geq 3$ for $1/2 \leq \gamma \leq 1$. Then with α any constant satisfying $0 < \alpha < (1/2)(\min(\gamma, 1/2) - 1/\ell)$, we have

(a) under (C1): For each fixed $t > 0$ and each fixed sequence of sampling rates $W_n \uparrow \infty$ as $n \uparrow \infty$, we have with probability one

$$(W_N)^\alpha \sup_{n \geq N} |\hat{s}_{W_n}(t) - s(t)| \rightarrow 0 \text{ as } N \rightarrow \infty ,$$

(b) under (C2): For each fixed $0 < t < 1$ and with $W = n$, a positive integer, we have with probability one

$$N^\alpha \sup_{n \geq N} |\hat{s}_n(t) - s(t)| \rightarrow 0 \text{ as } N \rightarrow \infty .$$

As an example, when $f(x)$ is bounded and monotonic (e.g. hardlimiter, quantizer) we have $\alpha < (1/2)\min(\gamma, 1/2)$ (as ℓ may be taken arbitrarily large); and thus for Lip 1 signals we have, in particular,

$$(W_n)^\alpha |\hat{s}_{W_n}(t) - s(t)| \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

with probability one for all $\alpha < 1/4$.

We finally show that, under certain conditions, the estimate $\hat{s}_W(t)$ is asymptotically normal and asymptotically independent at distinct times.

Theorem 3.4. Assume that $s(t)$ is Lip γ , $1/2 < \gamma \leq 1$, and that Assumptions (A), (B) and (C) are satisfied. In addition, assume that under (B1) or (B2) we have $f(x) \in L_p[\phi(x; \sigma) dx]$ for all $p \geq 1$, and under (B3) that the noise X is uniform over $[-c, c]$ with $c > b$. For $t \in I$ define

$$\tilde{s}_W(t) = B_W(t) [\hat{s}_W(t) - s(t)] ,$$

where

$$B_W(t) = \mu^{-1}[s(t)] \text{Var}^{-1/2}[\hat{m}_W(t)]$$

- (a) For each fixed t in the interior of I , $\tilde{s}_W(t)$ is asymptotically standard normal as $W \rightarrow \infty$. Bounds on the normalization factor $B_W(t)$ can be obtained from (35).
- (b) For the Szász operator in (C1) ($A(z) \equiv 1$) we have, in addition, that the values of the process $\{\tilde{s}_W(t), t > 0\}$ at distinct t 's are asymptotically independent as $W \rightarrow \infty$.

Some comments on Theorem 3.4. First, the theorem remains true if the statement " $s(t)$ is Lip γ , $1/2 < \gamma \leq 1$ " is replaced by " $\omega(s; \delta) = o(\delta^{1/2})$ as $\delta \rightarrow 0$ " (cf. the proof of Theorem 4.4). Second, part (b) of Theorem 3.4 remains true if the Szász operator is replaced by a generalized Szász operator for which $A(z)$ is a polynomial (cf. the proof of Proposition 4.1(b)). The question of asymptotic independence in the Bernstein case (C2) is open at present. Finally, the normalizing factor $B_W(t)$ will take a simple form if the exact rate of convergence of $\text{Var}[\hat{m}_W(t)]$ can be established. Specifically, we have obtained in Theorem 4.4 upper and lower bounds on $\text{Var}[\hat{m}_W(t)]$ of the form

$$0 < A_1(t)/\sqrt{W} \leq \text{Var}[\hat{m}_W(t)] \leq A_2(t)/\sqrt{W}$$

for some specified functions $A_i(t)$, $i = 1, 2$. When $s(t)$ is a constant, we find $A_1(t) \equiv A_2(t)$ in which case the rate of convergence of $\text{Var}[\hat{m}_W(t)]$ is exactly $1/\sqrt{W}$. If it can be established that this rate is valid for all signals $s(t)$ satisfying Assumption (A), we would then obtain

$$A(t) = \lim_{W \rightarrow \infty} W^{1/2} \text{Var}[\hat{m}_W(t)]$$

and the central limit theorem for $\hat{s}_W(t)$ could be stated in the more standard form: $W^{1/4}[\hat{s}_W(t) - s(t)]$ is asymptotically normal with mean zero and variance

$$A(t)/\{\mu[s(t)]\}^2.$$

A final comment in this section. For signals $s(t)$ defined on $[0, \infty)$ we always assumed uniform continuity of $s(t)$ over $[0, \infty)$ and obtained results valid on $(0, \infty)$ (uniformly on finite subintervals). For signals defined on $[0, \infty)$ that are continuous but not uniformly continuous, using the results of [8], we obtain results similar to those of Sections II and III valid over finite subintervals of $(0, \infty)$ (and expressed in terms of the modulus of continuity of $s(t)$ over each such subinterval). These results are of obvious interest but are not stated here explicitly to avoid overburdening the text.

IV. CONVERGENCE PROPERTIES OF THE ESTIMATE $\hat{m}_W(t)$

Let

$$m(t) = E[f(s(t) + X)] = \mu(s(t)), \quad t \in I$$

be the mean function of the output of the nonlinearity $f(x)$, where $\mu(s)$ is defined in (10). We establish the mean-square consistency, strong consistency, and a central limit theorem for $\hat{m}_W(t)$, given in (1), as an estimate of $m(t)$. These results, which are of independent value, are given in Part (b). In Part (a) we collect certain properties of the function $\mu(s)$. In order not to overburden the

text, the proofs of all the propositions are delegated to an Appendix.

(a) Properties of the Moment Function $\mu(s)$

For each $k = 1, 2, \dots$ define $\mu_k(s)$ by

$$\mu_k(s) = E[f^k(s+X)], \quad -\infty < s < \infty. \quad (14)$$

When X is $N(0, \sigma^2)$, $\mu_k(s)$, denoted by $\mu_{N,k}(s)$, is well defined whenever $f(x) \in L_{2k}[\phi(x; \sigma)dx]$ as follows by the inequality

$$|\mu_{N,k}(s)| \leq e^{s^2/2\sigma^2} \{E[f^{2k}(X)]\}^{1/2} \quad (15)$$

shown in [3]. The following properties of $\mu_{N,1}(s)$, denoted simply by $\mu_N(s)$, were shown in [3]. $\mu_N(s)$ is infinitely differentiable. If $f(x) \in L_2[\phi(x; \sigma)dx]$ is monotonic (not necessarily strictly monotonic) then $\mu_N(s)$ is strictly monotonic and

$$\mu'_N(s) > 0 \quad \text{for all } s.$$

If $f(x) \in L_2[\phi(x; \sigma)dx]$ is odd and has nonnegative Hermite coefficients $\{e_k\}_{k=0}^{\infty}$, then $\mu_N(s)$ is strictly monotonic with $\mu'_N(s) \geq e_1$ for all s and if, moreover, $e_1 > 0$ then

$$\mu'_N(s) \geq e_1 > 0 \quad \text{for all } s.$$

We shall need (and use) strictly positive lower bounds on $|\mu'_N(s)|$. Note that it is possible to have $\mu'_N(s) \rightarrow 0$ as $|s| \rightarrow \infty$ (e.g., if $|f(x)| \leq M$ and $\lim_{x \rightarrow \pm\infty} f(x) = \pm M$) and in such cases s would have to be limited to a bounded set of values.

When X is uniform over $[-b, b]$ and $f(x) = \text{sgn } x$, then $\mu_{U,k}(s)$ is clearly well defined for all s and all k and $\mu_{U,1}(s)$, which is denoted by $\mu_U(s)$, is given by (3b). Also

$$\mu'_U(s) = \frac{1}{b} > 0 \quad \text{for } |s| \leq b.$$

In proving a central limit theorem for $\hat{m}_W(t)$ and $\hat{s}_W(t)$ we shall need the following property:

$$\min_{|s| \leq b} \text{Var}[f(s+X)] > 0. \quad (16)$$

When X is $N(0, \sigma^2)$, $\text{Var}[f(s+X)] = \mu_{N,2}(s) - \mu_{N,1}^2(s)$ which is a continuous function of s . Thus, to show (16), it suffices to show that $\text{Var}[f(s+X)] \neq 0$ for all $-\infty < s < \infty$. Indeed, if for some s , $\text{Var}[f(s+X)] = 0$, then $f(s+x) = \text{Const}$ for almost all x with respect to the normal density $\phi(x; \sigma)$ and thus $f(y) = \text{Const}$ for almost all y , which contradicts our hypothesis that $f(x)$ is not a constant function. When X is uniform over $[-c, c]$ with $c > b$, and $f(x) = \text{sgn } x$, then (16) follows from

$$\min_{|s| \leq b} \text{Var}[\text{sgn}(s+X)] = 1 - \max_{|s| \leq b} \left(\frac{s}{c}\right)^2 = 1 - \left(\frac{b}{c}\right)^2 > 0.$$

Finally we shall use the following finite and nonzero constants whose existence under Assumption (B1) or (B2) follows from the above discussion and under (B3) is evident.

$$q = \begin{cases} \min_{|s| \leq c} \mu_N'(s) & , \text{ under (B1)} \\ \min_{|s| < \infty} \mu_N'(s) = e_1 & , \text{ under (B2)} \\ \min_{|s| \leq b} \mu_U'(s) & , \text{ under (B3)} \end{cases} \quad (17a)$$

$$Q = \max_{|s| \leq b} \mu'(s) \quad , \text{ under (B)} \quad (17b)$$

$$U_2 = \max_{|s| \leq b} \text{Var}[f(s+X)] \quad , \text{ under (B)} \quad (17c)$$

$$V_2 = \min_{|s| \leq b} \text{Var}[f(s+X)] \quad , \text{ under (B)} \quad (17d)$$

$$\begin{aligned} \Delta &= \min\{\mu(c) - \mu(b), \mu(-b) - \mu(-c)\} \quad , \text{ under (B1)} \\ &\geq q(c-b). \end{aligned} \quad (17e)$$

Note that $V_2 = 0$ under (B3) but $V_2 > 0$ under the modified (B3) where X is uniform over $[-c, c]$ with $c > b$. V_2 is used only under modified (B3) when needed.

(b) Convergence Properties of $\hat{m}_W(t)$

We begin by considering the mean-square error for a fixed $t \in I$,

$$E[\hat{m}_W(t) - m(t)]^2 = \text{Bias}^2[\hat{m}_W(t)] + \text{Var}[\hat{m}_W(t)] .$$

We have

$$\hat{m}_W(t) = \sum_k Z_{W,k} h_W(t,k) , \quad (18)$$

with $Z_{W,k} = f(s(k/W) + X_k)$. Since the Z 's are independent we have for all k ,

$$E[Z_{W,k}^2] \leq \sup_{|s| \leq b} E[f(s+X)]^2 < \infty$$

by Assumptions (A) and (B) (cf. (15)). Thus the series (18) converges in quadratic mean, as well as with probability one, provided $\sum_k h_W^2(t,k) < \infty$. Then, since $E[Z_{W,k}] = \mu(s(k/W)) = m(k/W)$, we have

$$E[\hat{m}_W(t)] = \sum_k m(k/W) h_W(t,k) \triangleq P_W(m,t) .$$

If $h_W(t,k) \geq 0$ for all k , then P_W is a positive linear operator and by a well-known result in approximation theory (see, for instance, Devore [9, pp. 28-29]), if $\sum_k h_W(t,k) = 1$ and $m(t)$ is uniformly continuous on I , then

$$|\text{Bias}[\hat{m}_W(t)]| = |P_W(m,t) - m(t)| \leq 2\omega(m; \alpha_W(t)) , \quad (19)$$

where $\omega(m; \delta)$ is the modulus of continuity of $m(t)$ over I and

$$\alpha_W^2(t) \triangleq P_W((\tau - t)^2, t) = \sum_k \left(\frac{k}{W} - t\right)^2 h_W(t,k) . \quad (20)$$

Also, using (17c), we have

$$\text{Var}[\hat{m}_W(t)] = \sum_k \text{Var}[Z_{W,k}] h_W^2(t,k) \leq U_2 v_W^2(t) \quad (21)$$

where

$$v_W^2(t) \triangleq \sum_k h_W^2(t,k) . \quad (22)$$

Hence for each $t \in I$ for which $h_W(t,k) \geq 0$ for all k , $\sum_k h_W(t,k) = 1$, and $\sum_k h_W^2(t,k) < \infty$, we have

$$E[\hat{m}_W(t) - m(t)]^2 \leq 4\omega^2(m; \alpha_W(t)) + U_2 v_W^2(t) \quad (23)$$

Thus if $\alpha_W^2(t) \rightarrow 0$ and $v_W^2(t) \rightarrow 0$ as $W \rightarrow \infty$ it follows that $\hat{m}_W(t)$ converges in quadratic mean to $m(t)$ as $W \rightarrow \infty$. This simple result is stated below.

Theorem 4.0. Under Assumptions (A) and (B), and for every $t \in I$ for which h_W satisfies conditions (i)-(iv) of Theorem 3.0, we have that $\hat{m}_W(t)$ converges in quadratic mean to $m(t)$ as $W \rightarrow \infty$.

The following proposition, whose proof is given in the Appendix, is used in determining bounds on the rate of convergence of $\hat{m}_W(t)$.

Proposition 4.1. (a) Let $\Psi(\lambda)$ be a 2π -periodic function continuously differentiable on $[-\pi, \pi]$ with Fourier series $\Psi(\lambda) = \sum_{k=-\infty}^{\infty} \psi_k \exp(ik\lambda)$. Then for $\ell = 2, 3, \dots$,

$$\sum_{k=-\infty}^{\infty} (\psi_k)^\ell = \frac{1}{(2\pi)^{\ell-1}} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \Psi \left(\sum_{i=1}^{\ell-1} \lambda_i \right) \prod_{j=1}^{\ell-1} [\Psi(\lambda_j) d\lambda_j] \quad .$$

(b) For the generalized Szász operator (13) we have for $t, t_1, t_2 > 0$,

$$(i) \quad \sum_{k=0}^{\infty} h_W(t_1, k) h_W(t_2, k) = \frac{\exp(-W(t_1 + t_2))}{2A^2(1)} \sum_{k=0}^{\infty} \epsilon_k b_k \left[\left(\frac{t_1}{t_2} \right)^{k/2} + \left(\frac{t_2}{t_1} \right)^{k/2} \right] I_k(2W \sqrt{t_1 t_2})$$

where $\epsilon_0 = 1$, $\epsilon_k = 2$ for $k \geq 1$, $b_k = \sum_{j=0}^{\infty} a_{k+j} a_j$, $k \geq 0$ and $I_k(x)$ is the modified Bessel function of the first kind of order k .

$$(ii) \quad v_W^2(t) \triangleq \sum_{k=0}^{\infty} h_W^2(t, k) = \frac{1 + o(1)}{2\sqrt{\pi W t}} \quad ,$$

$$\Gamma e^{-2Wt} I_0(2Wt) \leq v_W^2(t) \leq e^{-2Wt} I_0(2Wt) \quad ,$$

$$\text{where} \quad \Gamma = A^{-2}(1) \sum_{k=0}^{\infty} a_k^2 \quad .$$

$$(iii) \quad \sum_{k=0}^{\infty} [h_W(t, k)]^\ell \leq (e^{-Wt} I_0(Wt))^{\ell-1} = \left(\frac{1 + o(1)}{2\pi W t} \right)^{(\ell-1)/2} \quad , \quad \ell \geq 3 \quad .$$

(c) For the Bernstein operator (6) we have for $0 < t < 1$,

$$(i) \quad v_W^2(t) \triangleq \sum_{k=0}^W h_W^2(t, k) = \frac{1 + o(1)}{2\sqrt{\pi W t(1-t)}}.$$

$$(ii) \quad \sum_{k=0}^W [h_W(t, k)]^\ell \leq \left(\frac{1 + o(1)}{2\pi W t(1-t)} \right)^{(\ell-1)/2}, \quad \ell \geq 3.$$

Theorem 4.1. Under Assumptions (A), (B) and (C) we have for each $t \in I$,

$$E[\hat{m}_W(t) - m(t)]^2 \leq 4\omega^2(m; \alpha_W(t)) + U_2 v_W^2(t)$$

where the constant U_2 is given by (17c) and $\alpha_W(t)$ and $v_W(t)$ are as in Theorem 3.1.

Proof. The general bound on the mean-square error is given by (23). We only need to show that $m(t)$ is uniformly continuous on I and to compute $\alpha_W^2(t)$ and $v_W^2(t)$ under (C1) and (C2). Since $m(t) = \mu(s(t))$ and $\mu(s)$ is continuously differentiable with bounded derivative over the range of $s(t)$ (cf. 17(b)), the uniform continuity of $m(t)$ follows from that of $s(t)$. In fact it is easily seen that

$$\omega(m; \delta) \leq Q\omega(s; \delta) \quad (25)$$

where the constant Q is finite by (17b). Next we compute $\alpha_W^2(t)$ and $v_W^2(t)$ under (C1) and (C2). For the generalized Szász operators, (C1), we have by (20) and (13)

$$\alpha_W^2(t) = \frac{e^{-Wt}}{A(1)W^2} \sum_{k=0}^{\infty} (k-Wt)^2 p_k(Wt) = \frac{t}{W} + \frac{A''(1) + A'(1)}{W^2 A(1)}$$

where the last step follows by the expression for the series given in [6]. The expression for $v_W^2(t)$ under (C1) follows by Proposition 4.1(b.ii). For the Bernstein operator, (C2), $\alpha_W^2(t)$ is equal to the variance $t(1-t)/W$ of the binomial distribution (W, t) and $v_W^2(t)$ is given by Proposition 4.1(c). \square

Next we consider the convergence of $\hat{m}_W(t)$ in the $2\ell^{\text{th}}$ mean. The following proposition on the cumulants of $\hat{m}_W(t)$ is needed and its proof is given in the Appendix.

Proposition 4.2. Let Assumptions (A), (B), and (C) be satisfied. Let r be a positive integer and under (B1) or (B2) assume, in addition, that $f(x) \in L_{2r}[\phi(x;\sigma)dx]$. Then for every choice of points $\{t_1, \dots, t_r\}$ in I , the joint cumulant of $\hat{m}_W(t)$ of order r satisfies

$$(a) \quad |\text{Cum}_r \{\hat{m}_W(t_1), \dots, \hat{m}_W(t_r)\}| \leq M_r \sum_k \prod_{i=1}^r h_W(t_i, k)$$

for some finite positive constant M_r .

$$(b) \quad |\text{Cum}_r \{\hat{m}_W(t_1), \dots, \hat{m}_W(t_r)\}| \leq \frac{M_r [1 + o(1)]}{(2\pi W)^{(r-1)/2} \left\{ \prod_{i=1}^r D(t_i) \right\}^{(r-1)/2r}}$$

where

$$D(t) = \begin{cases} t & , \text{ under (C1)} \\ t(1-t) & , \text{ under (C2)} \end{cases} \quad (26)$$

Theorem 4.2. Under the Assumptions of Theorem 3.2 we have

$$E[\hat{m}_W(t) - m(t)]^{2\ell} \leq \frac{F_{\ell, \gamma}(t)}{W^{\ell \min(\gamma, 1/2)}} (1 + o(1))$$

for some continuous function $F_{\ell, \gamma}(t)$ specified in the proof.

Proof. For notational convenience we write \hat{m} , m for $\hat{m}_W(t)$, $m(t)$, respectively. Since $\hat{m} - m = \text{Bias}[\hat{m}] + (\hat{m} - E[\hat{m}])$, we have

$$E[\hat{m} - m]^{2\ell} = (\text{Bias}[\hat{m}])^{2\ell} + \sum_{j=2}^{2\ell} \binom{2\ell}{j} (\text{Bias}[\hat{m}])^{2\ell-j} E[\hat{m} - E[\hat{m}]]^j. \quad (27)$$

Since an estimate for $\text{Bias}[\hat{m}]$ has already been obtained in (19), we seek an estimate for $E[\hat{m} - E[\hat{m}]]^j$. We recall that with $n \triangleq \hat{m} - E[\hat{m}]$

$$E[n^r] = \sum_{p=1}^r \sum_{i=1}^p \prod_{v_i} \text{Cum}_{v_i} \{n, \dots, n\}, \quad r \geq 2, \quad (28)$$

where the inner sum extends over all partitions (v_1, \dots, v_p) of the set $\{1, \dots, r\}$

satisfying $v_1 + \dots + v_p = r$ [10]. Now any partition (v_1, \dots, v_p) with $p > [r/2]$ (the integer part of $r/2$), will necessarily have a factor $\text{Cum}_1\{n\} = E[n] \equiv 0$ in the product of cumulants in (28). Thus the range of p in (28) is reduced to $p = 1, \dots, [r/2]$. Next we note that $\text{Cum}_v\{n, \dots, n\} = \text{Cum}_v\{\hat{m}, \dots, \hat{m}\}$ for $v \geq 2$ and by Proposition 4.2(b) we have

$$|\text{Cum}_v\{n, \dots, n\}| \leq \frac{M_v[1 + o(1)]}{\{2\pi W D(t)\}^{(v-1)/2}}, \quad v \geq 2.$$

Thus for each $p=1, \dots, [r/2]$ we have

$$\left| \sum_{i=1}^p \prod_{j=1}^p \text{Cum}_{v_j}\{n, \dots, n\} \right| \leq H_p \frac{[1 + o(1)]}{\{2\pi W D(t)\}^{(r-p)/2}}, \quad (29)$$

with $H_p \triangleq \sum_{i=1}^p M_{v_i}$. (29) implies that the dominant term in (28) as $W \rightarrow \infty$ corresponds to $p = [r/2]$ so that for $r \geq 2$,

$$|E(\hat{m} - E[\hat{m}])^r| \leq H_{[r/2]} \frac{[1 + o(1)]}{\{2\pi W D(t)\}^{(r-[r/2])/2}}, \quad r \geq 2. \quad (30)$$

Since $s(t)$ is Lip γ , $0 < \gamma \leq 1$, i.e., $\omega(s; \delta) \leq L_s \delta^\gamma$, then by (25) $m(t)$ is also Lip γ with

$$\omega(m; \delta) \leq L_m \delta^\gamma; \quad L_m = L_s Q.$$

Thus from (19) and $\alpha_W^2(t) = (D(t)/W)[1 + o(1)]$ (cf. expressions in Theorem 3.1) we have

$$|\text{Bias}[\hat{m}]| \leq 2L_m (D(t)/W)^{\gamma/2} [1 + o(1)]. \quad (31)$$

It then follows by (30) and (31) that (27) can be bounded by

$$\begin{aligned} E[\hat{m} - m]^{2\ell} &\leq (2L_m)^{2\ell} (D(t)/W)^{\gamma\ell} [1 + o(1)] \\ &\quad + [1 + o(1)] \sum_{j=2}^{2\ell} \binom{2\ell}{j} (2L_m)^{2\ell-j} \frac{\{D(t)\}^{\gamma\ell - (j\gamma/2)} H_{[j/2]}}{\{2\pi D(t)\}^{(j-[j/2])/2}} \cdot \frac{1}{W^{\theta_j}} \end{aligned} \quad (32)$$

where

$$\theta_j = \ell\gamma - \frac{j}{2}(\gamma-1) - \frac{[j/2]}{2}, \quad j = 2, 3, \dots, 2\ell.$$

We now seek the dominant term in the above bound as $W \rightarrow \infty$. This depends on the value of γ .

(a) For $1/2 < \gamma \leq 1$, the dominant term in (32) corresponds to $j = 2\ell$ for which $\theta_{2\ell} = \ell/2$ and thus

$$E[\hat{m}-m]^{2\ell} \leq \frac{H_\ell[1 + o(1)]}{\{2\pi D(t)\}^{\ell/2} W^{\ell/2}}.$$

(b) For $0 < \gamma < 1/2$, the sum $\sum_{j=2}^{2\ell}$ is $o(W^{-\gamma\ell})$ so that

$$E[\hat{m}-m]^{2\ell} \leq \frac{(2L_m)^{2\ell} \{D(t)\}^{\gamma\ell}}{W^{\gamma\ell}} [1 + o(1)].$$

(c) For $\gamma = 1/2$, the terms in (32) corresponding to j odd are $o(W^{-\ell/2})$ and are negligible relative to the remaining terms. Thus

$$E[\hat{m}-m]^{2\ell} \leq \frac{\{D(t)\}^{\ell/2}}{W^{\ell/2}} \left\{ \sum_{\substack{j=0 \\ j \text{ even}}}^{2\ell} \binom{2\ell}{j} \frac{(2L_m)^{2\ell-j}}{\{2\pi D(t)\}^{j/2}} H_{j/2} \right\} [1 + o(1)].$$

These results can be combined for all $0 < \gamma \leq 1$ in the form given in the theorem where $F_{\ell,\gamma}(t)$ can easily be identified from the above analysis. \square

We next obtain the strong consistency of $\hat{m}_W(t)$. The result is identical to Theorem 3.3 but with $\hat{m}_W(t)-m(t)$ replacing $\hat{s}_W(t)-s(t)$.

Theorem 4.3. Under the Assumptions of Theorem 3.3, $\hat{m}_W(t)-m(t)$ satisfies its conclusion.

Proof. Fix t in the interior of I and consider the estimate $\hat{m}_W(t)$ as a function of W : Define a process $\{\eta_u, u \geq 0\}$ by

$$\eta_u = \begin{cases} m(t) & , \quad u = 0 \\ \hat{m}_{1/u}(t) & , \quad u > 0 \end{cases} \quad (33)$$

(a) Under (C1), $\{\eta_u, u \geq 0\}$ is not necessarily separable. Fix a sequence $\{W_n\}_{n=1}^\infty$ with $W_n \uparrow \infty$ and let $\{\tilde{\eta}_u, u \geq 0\}$ be a separable version of $\{\eta_u, u \geq 0\}$ with a separating set which includes the points $u_0 = 0$ and $u_n = 1/W_n, n \geq 1$. Then for any $\delta > 0$ we have

$$\sup_{u_n \leq \delta} |\eta_{u_n} - \eta_0| \leq \sup_{u \leq \delta} |\tilde{\eta}_u - \tilde{\eta}_0| \quad (34)$$

Now, since the two processes $\{\eta_u, u \geq 0\}$ and $\{\tilde{\eta}_u, u \geq 0\}$ have the same finite dimensional distributions, it follows by Theorem 4.2 that

$$E|\tilde{\eta}_u - \tilde{\eta}_0|^{2\ell} \leq \{K_{\ell, \gamma}(t)[1 + o(1)]\} u^{\beta+1},$$

where $\beta = \ell \min(\gamma, 1/2) - 1$. It then follows by Kolmogorov's theorem (see Neveu [11, p. 97]) that with probability one

$$\frac{1}{\delta^\alpha} \sup_{u \leq \delta} |\tilde{\eta}_u - \tilde{\eta}_0| \rightarrow 0 \quad \text{as } \delta \downarrow 0,$$

for any $0 < \alpha < \beta/2\ell$. Hence by (33) and (34) we have, with probability one,

$$\frac{1}{\delta^\alpha} \sup_{1/W_n \leq \delta} |\hat{m}_{W_n}(t) - m(t)| \rightarrow 0 \quad \text{as } \delta \downarrow 0,$$

and the result follows by choosing $\delta = 1/W_n$.

(b) Under (C2), $W = n$ (an integer) so that $\{\eta_u, u \geq 0\}$ is separable. Theorem 4.2 and Kolmogorov's theorem imply that, with probability one,

$$\frac{1}{\delta^\alpha} \sup_{1/n \leq \delta} |\eta_{1/n} - \eta_0| \rightarrow 0 \quad \text{as } \delta \downarrow 0,$$

and the result follows by choosing $\delta = 1/N$. \square

We finally derive a central limit theorem for the estimate $\hat{m}_W(t)$.

Define the normalized error process

$$\tilde{m}_W(t) = \frac{\hat{m}_W(t) - m(t)}{\text{Var}^{1/2}[\hat{m}_W(t)]}, \quad t \in I.$$

Theorem 4.4. Under the assumptions of Theorem 3.4 we have

(a) For each fixed t in the interior of I , $\tilde{m}_W(t)$ is asymptotically standard normal variable as $W \rightarrow \infty$. The normalizing factor $\text{Var}^{-1/2}[\hat{m}_W(t)]$ satisfy

$$\begin{aligned} \{2\sqrt{\pi D(t)}/U_2\}^{1/2} W^{1/4}[1 + o(1)] &\leq \text{Var}^{-1/2}[\hat{m}_W(t)] \\ &\leq \{2\sqrt{\pi D(t)}/V_2\}^{1/2} W^{1/4}[1 + o(1)] \end{aligned} \quad (35)$$

where the constants U_2 and V_2 are given in (17) and $D(t)$ is given by (26).

(b) For the Szász operator in (C1) ($A(z) \equiv 1$), we have in addition, that the values of the process $\{\tilde{m}_W(t), t > 0\}$ at distinct t 's are asymptotically independent as $W \rightarrow \infty$.

Proof. Putting

$$\xi_W(t) = \frac{\hat{m}_W(t) - E[\hat{m}_W(t)]}{\text{Var}^{1/2}[\hat{m}_W(t)]},$$

we have that

$$\tilde{m}_W(t) = \xi_W(t) + \frac{\text{Bias}[\hat{m}_W(t)]}{\text{Var}^{1/2}[\hat{m}_W(t)]}.$$

The proof is accomplished by showing that as $W \rightarrow \infty$ the second term goes to zero and $\xi_W(t)$ has the asymptotic properties stated in the theorem.

Under (B1), (B2) and the modified (B3), we have by (17c) - (17d)

$$0 < V_2 \leq \text{Var}[Z_{W,k}] \leq U_2 < \infty, \quad (36)$$

and thus by (21),

$$V_2 v_W^2(t) \leq \text{Var}[\hat{m}_W(t)] \leq U_2 v_W^2(t) \quad (37)$$

(with equality when $s(t)$ is constant). Using the asymptotic expression for $v_W^2(t)$ given in Proposition 4.1(b)(c) we have

$$\frac{v_2[1 + o(1)]}{2\sqrt{\pi D(t)W}} \leq \text{Var}[\hat{m}_W(t)] \leq \frac{u_2[1 + o(1)]}{2\sqrt{\pi D(t)W}} . \quad (38)$$

Hence by (31) and (38), since $\gamma > 1/2$,

$$\frac{|\text{Bias}[\hat{m}_W(t)]|}{\text{Var}^{1/2}[\hat{m}_W(t)]} = o(W^{-1/2(\gamma-1/2)}) \rightarrow 0 .$$

We now establish the desired asymptotic results for $\xi_W(t)$ for t in the interior of I . It is clear that

$$E[\xi_W(t)] = 0, \text{Var}[\xi_W(t)] = 1 .$$

For Part (a) we show that for each fixed t , all cumulants of $\xi_W(t)$ of order $r \geq 3$ tend to zero as $W \rightarrow \infty$; the asymptotic normality of $\xi_W(t)$ follows then from Lemma P4.5 of [12]. For Part (b) we show that for all $r \geq 3$ and all instants $t_1, \dots, t_r > 0$, not necessarily distinct, the joint cumulant

$$\text{Cum}_r\{\xi_W(t_1), \dots, \xi_W(t_r)\} \rightarrow 0 \text{ as } W \rightarrow \infty , \quad (39)$$

and in addition

$$E[\xi_W(t_1)\xi_W(t_2)] \rightarrow 0 \text{ as } W \rightarrow \infty \text{ for } t_1 \neq t_2 . \quad (40)$$

It will then follow by the same Lemma of [12] that all finite dimensional distributions of the process $\{\xi_W(t), t > 0\}$ converge to the finite dimensional distributions of a Gaussian process with mean zero and covariance $R(t_1, t_2) = 1$ for $t_1 = t_2$, and $R(t_1, t_2) = 0$ for $t_1 \neq t_2$, i.e., with independent values at distinct points. Both goals will be achieved if we show (39) in general, and (40) in the Szász case, which we now proceed to do. For $r \geq 3$ and $\{t_i\}$ in the interior of I , we have

$$\text{Cum}_r\{\xi_W(t_1), \dots, \xi_W(t_r)\} = \frac{\text{Cum}_r\{\hat{m}_W(t_1), \dots, \hat{m}_W(t_r)\}}{\prod_{i=1}^r \text{Var}^{1/2}[\hat{m}_W(t_i)]}$$

and using the upper bound in Proposition 4.2(b) for the numerator and the lower bound in (38) for each factor in the denominator, we obtain

$$|\text{Cum}_r\{\varepsilon_W(t_1), \dots, \varepsilon_W(t_r)\}| = o(W^{-(r-2)/4}) \rightarrow 0$$

since $r \geq 3$. Next we prove (40) for the special case of the Szász operator in (C1).

Note that by Proposition 4.1(b), specialized for the Szász case, we have

$$\sum_{k=0}^{\infty} h_W(t_1, k) h_W(t_2, k) = e^{-W(t_1+t_2)} I_0(2W\sqrt{t_1 t_2}) \quad (41)$$

Now

$$\text{Cov}\{\hat{m}_W(t_1), \hat{m}_W(t_2)\} = \sum_{k=0}^{\infty} \text{Var}[Z_{W,k}] h_W(t_1, k) h_W(t_2, k)$$

and by (36) and (41)

$$|\text{Cov}\{\hat{m}_W(t_1), \hat{m}_W(t_2)\}| \leq U_2 e^{-W(t_1+t_2)} I_0(2W\sqrt{t_1 t_2})$$

By (37) and (41) we have $\text{Var}[\hat{m}_W(t)] \geq V_2 e^{-2Wt} I_0(2Wt)$, so that

$$|E[\varepsilon_W(t_1) \varepsilon_W(t_2)]| \leq \frac{U_2}{V_2} \frac{I_0(2W\sqrt{t_1 t_2})}{\{I_0(2Wt_1) I_0(2Wt_2)\}^{1/2}}$$

Using the asymptotic expansion [13, p. 86] for large x , $I_0(x) = (2\pi x)^{-1/2} e^x (1 + o(1/x))$, we obtain for $t_1 \neq t_2$ as $W \rightarrow \infty$,

$$|E[\varepsilon_W(t_1) \varepsilon_W(t_2)]| \leq \frac{U_2}{V_2} e^{-W(\sqrt{t_1} - \sqrt{t_2})^2} [1 + o(1)] \rightarrow 0$$

Finally, the bounds on $\text{Var}^{-1/2}[\hat{m}_W(t)]$ follow from (37) and Proposition 4.1. \square

V. PROOFS OF THEOREMS OF SECTIONS II AND III

Using the convergence results for $\hat{m}_W(t)$, proven in Section IV(b), and the relationships $\hat{s}_W(t) = g[\hat{m}_W(t)]$, $m(t) = \mu[s(t)]$, we now establish the convergence results for $\hat{s}_W(t)$ stated in Sections II and III. The basic link between the properties of $\hat{s}_W(t)$ and $\hat{m}_W(t)$ is provided by the following proposition whose proof is given in the Appendix.

Proposition 5.1. Let Assumptions (A) and (C) be satisfied. Then

(a) under (B1), with $p \geq 1$, we have

$$E|\hat{s}_W(t) - s(t)|^p \leq [(1/q)^p + (b/\Delta)^p] E|\hat{m}_W(t) - m(t)|^p,$$

(b) under (B2) we have

$$|\hat{s}_W(t) - s(t)| \leq \left(\frac{1}{q}\right) |\hat{m}_W(t) - m(t)|,$$

(c) under (B3) we have

$$\hat{s}_W(t) - s(t) = b[\hat{m}_W(t) - m(t)],$$

where the constants q and Δ are defined in (17) and b is the upper bound for $s(t)$.

Theorems 3.0-3.2 follow immediately from Theorems 4.0-4.2, respectively, and Proposition 5.1. Theorem 3.3 follows from Theorem 3.2 and Kolmogorov's theorem [11, p. 97] in the manner of the proof of Theorem 4.3. The deduction of Theorem 3.4 from Theorem 4.4 is given below. Finally, Theorems 2.1-2.4 follow immediately from Theorems 3.1-3.4, respectively. (In Theorem 2.1, for the estimate (8a) under (B1), the values of the constants K_1 , K_2 are obtained from those of Theorem 3.1 by using the computed values $q = 2\phi(b+\epsilon, \sigma)$, $Q = \sqrt{2/\pi}/\sigma$, $U_2 = 1$, and the inequality $\Delta \geq q(c-b)$; the use of this inequality results in a simple expression for K_1 and K_2 .)

Proof of Theorem 3.4. (a) Fix t in the interior of the interval I . By Theorem 4.4(a), the distribution of $[\hat{m}_W(t) - m(t)]/\text{Var}^{1/2}[\hat{m}_W(t)]$ converges to the distribution of a standard normal variable, say z_t . A result of Mann and Wald [14, p. 226] shows that if $g(x)$ has a continuous first derivative in the neighborhood of $m(t)$, and $g'(m(t)) \neq 0$, then the distribution of $\{g[\hat{m}(t)] - g[m(t)]\}/\text{Var}^{1/2}[\hat{m}_W(t)]$ converges to the distribution of the normal variable $g(m(t))z_t$. Since $|s(t)| \leq b$, $m(t) = \mu(s(t))$ takes values in the interval $[\mu(-b), \mu(b)]$ for all $t \in I$. Thus under (B1), (B2) or the modified (B3) stated in the theorem, $g(x)$ is continuously differentiable over an interval containing $[\mu(-b), \mu(b)]$ and $g'(x) > 0$ for $\mu(-b) \leq x \leq \mu(b)$. It follows that the distribution of

$$\tilde{s}_W(t) \triangleq \frac{\hat{s}_W(t) - s(t)}{\text{Var}^{1/2}[\hat{m}_W(t)]} = \frac{g[\hat{m}_W(t)] - g[m(t)]}{\text{Var}^{1/2}[\hat{m}_W(t)]}$$

converges to the distribution of the normal variable $g'(m(t))z_t$ and the result follows from $g'(m(t)) = 1/\mu'[s(t)]$.

(b) Let $\{t_i\}_{i=1}^k$ be distinct points in $(0, \infty)$. By Theorem 4.4(b), the distribution of $\{\hat{m}_W(t_i)\}_{i=1}^k$ converges to the distribution of independent standard normal variables, say $\{z_i\}_{i=1}^k$, as $W \rightarrow \infty$. Again by the result of Mann and Wald [14, p. 226], the distribution of $\sum_{i=1}^k \theta_i \tilde{s}_W(t_i)$ converges to the distribution of the normal variable $\sum_{i=1}^k \theta_i g'[m(t_i)]z_i$ whose mean is zero and variance is $\sum_{i=1}^k \theta_i^2 \{g'[m(t_i)]\}^2$. Since the θ_i 's are arbitrary, it follows that the variables $\{\tilde{s}_W(t_i)\}_{i=1}^k$ are asymptotically independent normal. \square

APPENDIX

A. Proof of Proposition 3.0. It is clear by (11) that $h_n(t, k) \geq 0$ for all k and $\sum_k h_n(t, k) = 1$. Also, that $z_n = \xi_1 + \dots + \xi_n$ has mean nt and variance $n \text{Var}[\xi_1]$. Hence

$$\sum_{k=-\infty}^{\infty} \left(\frac{k}{n} - t \right)^2 h_n(t, k) = \frac{1}{n^2} \text{Var}[z_n] = \frac{1}{n} \text{Var}[\xi_1] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Thus conditions (i)-(iii) of Theorem 3.0 are satisfied for all $t \in I$. For (iv) we have

$$\begin{aligned} \sum_{k=-\infty}^{\infty} h_n^2(t, k) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |\phi_{z_n}(\lambda)|^2 d\lambda \\ &= \lim_{T \rightarrow \infty} \frac{n}{2T} \int_{-T/n}^{T/n} |\phi_{\xi}(\lambda)|^{2n} d\lambda \end{aligned}$$

where $\phi_{z_n}(\lambda)$, $\phi_{\xi}(\lambda)$ is the characteristic function of z_n , ξ_1 , respectively. Since ξ_1 is integer valued, $\phi_{\xi}(\lambda)$ is periodic with period 2π . Consider all $t \in I$ for which ξ_1 is a nondegenerate random variable. Then $\phi_{\xi}(\lambda)$ has a positive fundamental period which, without loss of generality, can be taken as 2π . Then

$$\sum_{k=-\infty}^{\infty} h_n^2(t, k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_{\xi}(\lambda)|^{2n} d\lambda$$

and (iv) follows by dominated convergence since $|\phi_{\xi}(\lambda)| < 1$ for $0 < |\lambda| < \pi$. \square

B. Proof of Proposition 4.1. (a) We have $\psi_k = (1/2\pi) \int_{-\pi}^{\pi} \psi(\lambda) \exp(-ik\lambda) d\lambda$ and thus

$$\begin{aligned} \sum_{|k| \leq N} (\psi_k)^{\ell} &= \sum_{|k| \leq N} \psi_k \frac{1}{(2\pi)^{\ell-1}} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \prod_{j=1}^{\ell-1} [\psi(\lambda_j) e^{-ik\lambda_j} d\lambda_j] \\ &= \frac{1}{(2\pi)^{\ell-1}} \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} \psi_N \left(- \sum_{i=1}^{\ell-1} \lambda_i \right) \prod_{j=1}^{\ell-1} [\psi(\lambda_j) d\lambda_j] \end{aligned}$$

where $\psi_N(\lambda) = \sum_{|k| \leq N} \psi_k e^{ik\lambda}$. Since $\psi(\lambda)$ is continuously differentiable,

$\psi_N(\lambda) \rightarrow \psi(\lambda)$ uniformly on $[-\pi, \pi]$ and, in fact, $\max_{|\lambda| \leq \pi} |\psi(\lambda) - \psi_N(\lambda)| \leq \text{Const } N^{-1/2}$ (see [15, p. 31]). The result follows by applying the dominated convergence theorem.

(b) By choosing $\psi_t(\lambda) = \exp[Wt(e^{i\lambda} - 1)] A(e^{i\lambda}) A^{-1}(1)$, we obtain from the generating function $A(z)$ of the Appel polynomials (12) that $\psi_k(t) \equiv h_W(t, k)$. Hence by Parseval's relationship

$$\begin{aligned} \sum_{k=0}^{\infty} h_W(t_1, k) h_W(t_2, k) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_{t_1}(\lambda) \psi_{t_2}^*(\lambda) d\lambda \\ &= \frac{e^{-W(t_1+t_2)}}{A^2(1)} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp[W(t_1 e^{i\lambda} + t_2 e^{-i\lambda})] |A(e^{i\lambda})|^2 d\lambda \right\}. \end{aligned}$$

But the Fourier series $A(e^{i\lambda}) = \sum_{k=0}^{\infty} a_k e^{ik\lambda}$ converges boundedly and uniformly on $[-\pi, \pi]$, since $A(z)$ is analytic in $|z| < R$ for some $R > 1$, so that by interchanging summation and integration (as in Part (a)) the expression in braces becomes

$$\begin{aligned} \sum_{j, \ell=0}^{\infty} a_j a_{\ell} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{W(t_1+t_2)\cos\lambda} e^{i[(j-\ell)\lambda + W(t_1-t_2)\sin\lambda]} d\lambda \\ = \sum_{j, \ell=0}^{\infty} a_j a_{\ell} (t_1/t_2)^{(\ell-j)/2} I_{\ell-j}(2W\sqrt{t_1 t_2}) \end{aligned}$$

by [16, p.488]. Noting that $I_n(x) = I_{-n}(x)$, and considering the sum for $j > \ell$ and $j < \ell$, we obtain Part (b.i). For (b.ii) we have from (b.i) with $t_1 = t_2 = t$ that

$$\begin{aligned} \sum_{k=0}^{\infty} h_W^2(t, k) &= \frac{e^{-2Wt}}{A^2(1)} \sum_{k=0}^{\infty} \epsilon_k b_k I_k(2Wt) \\ &\geq \frac{e^{-2Wt}}{A^2(1)} \epsilon_0 b_0 I_0(2Wt) = e^{-2Wt} I_0(2Wt) \end{aligned} \quad (A1)$$

On the other hand, since $|\psi_t(\lambda)| \leq \exp(-Wt)\exp(Wt \cos \lambda)$ we have

$$\sum_{k=0}^{\infty} h_W^2(t, k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi_t(\lambda)|^2 d\lambda \leq e^{-2Wt} I_0(2Wt),$$

which completes the proof for the upper and lower bounds in (b.ii). In order to obtain the asymptotic result in (b.ii) we note that [13, p. 86] as $x \rightarrow \infty$

$$I_n(x) = \frac{e^x}{\sqrt{2\pi x}} \left[1 + (4n^2 - 1) o\left(\frac{1}{x}\right) \right] \quad (A2)$$

where the term $o(1/x)$ is uniform in n . Hence as $W \rightarrow \infty$ we have by (A1)

$$\begin{aligned} \sum_{k=0}^{\infty} h_W^2(t, k) &= \frac{1}{A^2(1)\sqrt{4\pi Wt}} \left\{ \sum_{k=0}^{\infty} \epsilon_k b_k \left[1 + (4k^2 - 1) o\left(\frac{1}{2Wt}\right) \right] \right\} \\ &= \frac{1}{\sqrt{4\pi Wt}} \left\{ 1 + A^{-2}(1) o\left(\frac{1}{2Wt}\right) \sum_{k=0}^{\infty} (4k^2 - 1) \epsilon_k b_k \right\}, \end{aligned}$$

since $\sum_{k=0}^{\infty} \epsilon_k b_k = \left[\sum_{j=0}^{\infty} a_j \right]^2 = A^2(1)$. The asymptotic result will follow by showing $\sum_{k=0}^{\infty} (4k^2 - 1) \epsilon_k b_k < \infty$. Since $A(z) = \sum_{n=0}^{\infty} a_n z^n$ is analytic in $|z| < R$ for some $R > 1$, there exists a constant $0 < r < 1$ such that $a_n \leq \text{Const } r^n$. This implies that $b_n = \sum_{k=0}^{\infty} a_{n+k} a_k \leq \text{Const } r^n$. Thus $\sum_{k=0}^{\infty} k^2 b_k < \infty$ and the result follows. For Part (b.iii) we have by Part (a) with $\ell \geq 3$

$$\sum_{k=0}^{\infty} [h_W(t, k)]^{\ell} \leq \left\{ \max_{\lambda} |\psi_t(\lambda)| \right\}^{\ell} \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi_t(\lambda)| d\lambda \right\}^{\ell-1}$$

and the result follows by using the bound $|\psi_t(\lambda)| \leq \exp(-Wt)\exp(Wt \cos \lambda)$ and (A2):

$$\sum_{k=0}^{\infty} [h_W(t,k)]^L \leq [e^{-Wt} I_0(Wt)]^{L-1} = \left(\frac{1 + o(1)}{\sqrt{2\pi Wt}} \right)^{L-1}.$$

(c) By choosing $\psi_t(\lambda) = (t e^{i\lambda} + 1-t)^W$, which is the characteristic function of the binomial distribution, we have $\psi_k(t) = h_W(t,k)$ where $h_W(t,k)$ is given in (6). Part (c.i), (c.ii) follow in the manner of (b.ii), (b.iii), respectively, using the property

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} [1 - a \sin^2 \lambda/2]^W d\lambda = \frac{1 + o(1)}{\sqrt{\pi a W}}$$

for $0 < a < 1$, shown in [4]. \square

C. Proof of Proposition 4.2. (a) From (18) we have

$$\begin{aligned} \text{Cum}(\hat{m}_W(t_1), \dots, \hat{m}_W(t_r)) &= \sum_{k_1} \dots \sum_{k_r} \text{Cum}_r \{Z_{W,k_1}, \dots, Z_{W,k_r}\} \prod_{i=1}^r h_W(t_i, k_i) \\ &= \sum_k \text{Cum}_r \{Z_{W,k}, \dots, Z_{W,k}\} \prod_{i=1}^r h_W(t_i, k) \end{aligned} \quad (A3)$$

where the first equality is justified below and the second equality follows from the independence of $\{Z_{W,k}\}_k$ and the fact that the joint cumulant of independent sets of random variables is zero [12, p. 19]. Part (a) will follow from (A3) provided $|\text{Cum}_r \{Z_{W,k}, \dots, Z_{W,k}\}| \leq M_r$ for some finite positive constant M_r - which is seen as follows:

$$\text{Cum}_r \{Z_{W,k}, \dots, Z_{W,k}\} = \sum_{p=1}^r \sum (-1)^{p(p-1)!} \prod_{i=1}^p E[(Z_{W,k})^{v_i}] \quad (A4)$$

where the inner sum extends over all partitions (v_1, \dots, v_p) of the set $\{1, \dots, r\}$ satisfying $v_1 + \dots + v_p = r$ [12, p. 19]. By Assumption (A), for all k

$$E|Z_{W,k}|^v \leq \sup_{|s| \leq b} E|f(s+X)|^v = \text{Const}_v < \infty, \quad v = 1, \dots, r \quad (A5)$$

where the last step follows from (15) and $f(x) \in L_{2r}[\phi(x;\sigma)dx]$ under (B1) and (B2) (under (B3) this is obvious). Putting (A5) in (A4) gives the required bound M_r .

The first equality in (A3) is justified as follows. Since cumulants and moments can be expressed in terms of each other (cf. (A4) and (28)), it suffices to justify the exchange of expectation and summation for moments. This will follow by Fubini's theorem provided

$$\sum_{k_1} \cdots \sum_{k_r} E \left| \prod_{i=1}^r Z_{W,k_i} h_W(t_i, k_i) \right| \quad (A6)$$

is finite. But by the multi-dimensional version of Hölder's inequality, (A6) is bounded by

$$\sum_{k_1} \cdots \sum_{k_r} \prod_{i=1}^r \{E|Z_{W,k_i} h_W(t_i, k_i)|^r\}^{1/r} = \prod_{i=1}^r \left[\sum_{k_i} E|Z_{W,k_i}|^r h_W^r(t_i, k_i) \right]^{1/r}.$$

The latter is finite since $E|Z_{W,k}|^r < \infty$ by (A5) and $\sum_k [h_W(t, k)]^r < \infty$ by Proposition 4.1(b)(c).

(b) By the r^{th} dimensional version of Hölder's inequality for sums we have

$$\sum_k \prod_{i=1}^r h_W(t_i, k) \leq \prod_{i=1}^r \left\{ \sum_k [h_W(t_i, k)]^r \right\}^{1/r}$$

and the result follows by Proposition 4.1(b)(c). \square

D. Proof of Proposition 5.1

We provide the proofs in reverse order.

(c) Under (B3) we have $m(t) = (1/b)s(t)$ since $|s(t)| \leq b$. Also

$$\hat{m}_W(t) = \sum_k \text{sgn}[s(k/W) + x_k] h_W(t, k)$$

satisfies $|\hat{m}_W(t)| \leq \sum_k h_W(t, k) = 1$. Hence $g(x)$, given under (B3), is used only for $|x| \leq 1$ and thus $\hat{s}_W(t) = b \hat{m}_W(t)$.

(b) Under (B2), $\nu^{-1}(x)$ exists for all x , $s(t) = \nu^{-1}[m(t)]$ and $\hat{s}_W(t) = \nu^{-1}[\hat{m}_W(t)]$.

The result follows from the inequality

$$|\nu^{-1}(x) - \nu^{-1}(y)| \leq \frac{|x-y|}{\min_{|s| < \infty} \nu'(s)} = q|x-y|$$

which is valid for all $-\infty < x, y < \infty$.

(a) Under (B1) $\mu^{-1}(x)$ exists for all x and $s(t) = \mu^{-1}[m(t)]$. For simplicity we omit W and t in the following. We have by (B1)

$$|\hat{s}-s| = \begin{cases} |\mu^{-1}(\hat{m}) - \mu^{-1}(m)|, & \text{if } \mu(-c) \leq \hat{m} \leq \mu(c) \\ |s| & , \text{ otherwise.} \end{cases}$$

Also for $\mu(-c) \leq x, y \leq \mu(c)$,

$$|\mu^{-1}(x) - \mu^{-1}(y)| \leq \frac{|x-y|}{\min_{|s| \leq c} \mu'(s)} = q|x-y|$$

and thus

$$E|\hat{s}-s|^p \leq (1/q)^p E|\hat{m}-m|^p + |s|^p \Pr\{\hat{m} \notin [\mu(-c), \mu(c)]\}.$$

Now

$$\begin{aligned} \Pr\{\hat{m} \notin [\mu(-c), \mu(c)]\} &= 1 - \Pr\{\mu(-c) \leq \hat{m} \leq \mu(c)\} \\ &= 1 - \Pr\{\mu(-c)-m \leq \hat{m}-m \leq \mu(c)-m\} \\ &\leq 1 - \Pr\{|\hat{m}-m| \leq \Delta\} = \Pr\{|\hat{m}-m| > \Delta\} \\ &\leq (1/\Delta)^p E|\hat{m}-m|^p \end{aligned}$$

where the first inequality above follows from (17e) since $m(t) = \mu[s(t)] \Rightarrow \mu(-b) \leq m \leq \mu(b) \Rightarrow \mu(c)-m \geq \mu(c)-\mu(b) \geq 0$ and $\mu(-c)-m \leq \mu(-c)-\mu(-b) \leq 0$. The result follows. \square

References

- [1] E. C. Titchmarsh, "The zeros of certain integral functions," Proc. London Math. Soc., Vol. 25, pp. 283-302, 1925.
- [2] R. Duffin and A. C. Schaeffer, "Some properties of functions of exponential type," Bull. Amer. Math. Soc., Vol. 44, pp. 236-240, 1938.
- [3] E. Masry and S. Cambanis, "Signal identification after noisy nonlinear transformations," IEEE Trans. Inform. Theory, Vol. IT-26, Jan. 1980, to appear.
- [4] A. Ja. Dorogovcev, "On a nonparametric estimate of the regression function," Theory Probability Math. Statist., Vol. 4, pp. 55-57, 1974.
- [5] D. D. Stancu, "Use of probabilistic methods in the theory of uniform approximation of continuous functions," Rev. Roumaine. Math. Pures Appl., Vol. 14, pp. 673-691, 1969.
- [6] A. Jakimowski and D. Leviatan, "Generalized Szász operators for the approximation in the infinite interval," Mathematica, Vol. 11, pp. 97-103, 1969.
- [7] R. P. Boas and R. C. Buck, Polynomial Expansions of Analytic Functions, Berlin: Springer-Verlag, 1958.
- [8] E. Masry, "On the approximation of continuous functions by the generalized Szász operators," submitted.
- [9] R. A. Devore, The Approximation of Continuous Functions by Positive Linear Operators, Berlin: Springer-Verlag, 1972.
- [10] M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. I., London: Griffin, 1963.
- [11] J. Neveu, Mathematical Foundations of the Calculus of Probability, San Francisco: Holden-Day, 1965.

- [12] D. R. Brillinger, Time Series Data Analysis and Theory, New York: Holt, Rinehart and Winston, 1975.
- [13] A. Erdélyi, Ed., Higher Transcendental Functions, Vol. II, New York: McGraw-Hill, 1953.
- [14] H. B. Mann and A. Wald, "On stochastic limit and order relationships," Ann. Math. Statist., Vol. 14, pp. 217-226, 1943.
- [15] H. Dym and H. P. McKean, Fourier Series and Integrals, New York: Academic Press, 1972.
- [16] I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products, New York: Academic Press, 1965.

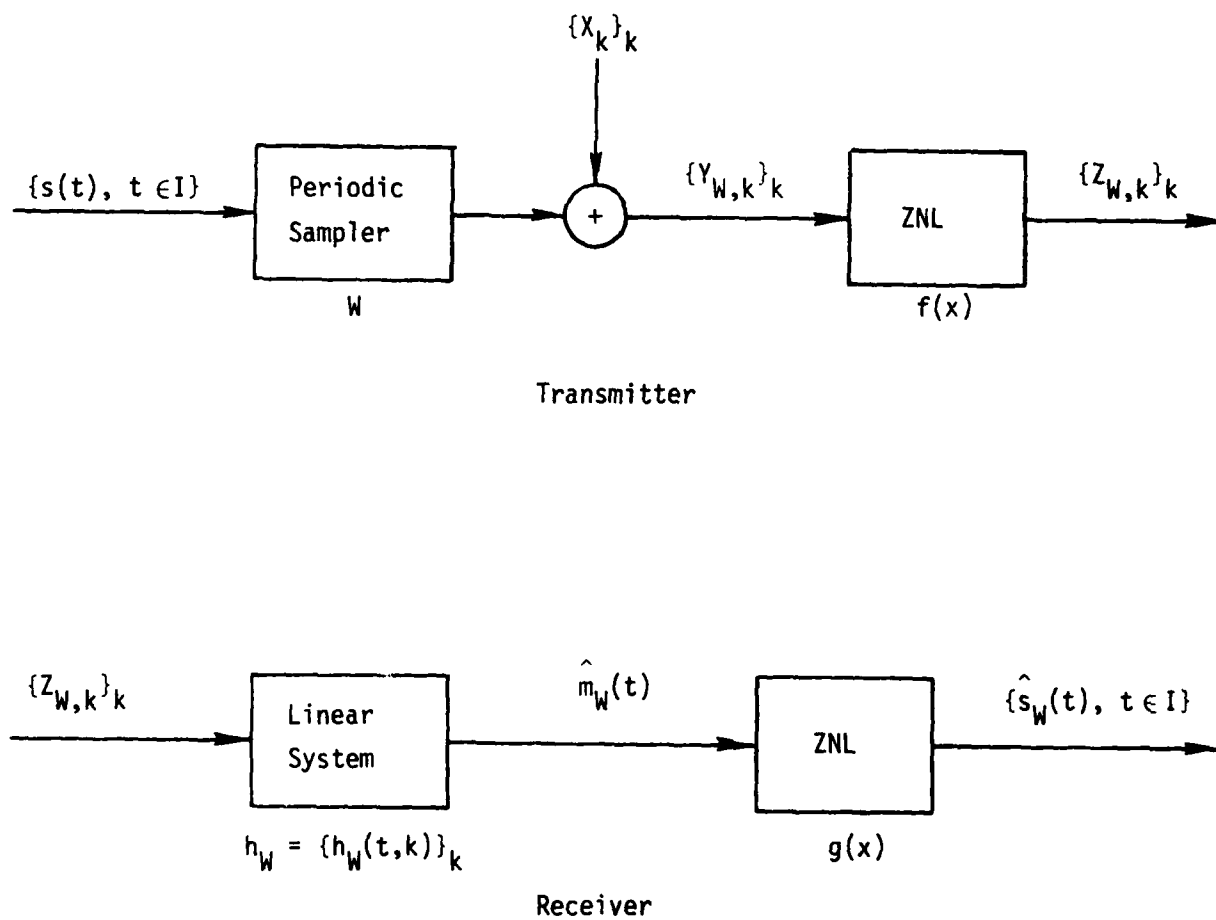


Figure 1. The structure of the transmitter/receiver model.